Mengulas Penggunaan Koefisien Alpha dalam Mengevaluasi Pengukuran Psikologi

Wahyu Widhiarso

Fakultas Psikologi UGM

2013

Permasalahan yang sering muncul dalam analisis data untuk mengevaluasi properti psikometri alat ukur di psikologi adalah penggunaan koefisien reliabilitas secara monoton tanpa mempertimbangkan asumsi yang mendasari koefisien tersebut. Socan (2000) mengatakan bahwa banyak peneliti yang terpaku pada penggunaan koefisien alpha (Cronbach, 1951). Banyak juga diantara peneliti yang tidak menyadari bahwa koefisien alpha menghendaki asumsi tertentu untuk dipenuhi. Jika asumsi ini tidak dipenuhi maka koefisien reliabilitas yang dihasilkan menjadi bias. Pemilihan formula reliabilitas secara monoton tersebut dapat diakibatkan oleh dua sebab: minimnya pemahaman mengenai koefisien reliabilitas alternatif dan minimnya keberadaan program komputasi yang dapat mengelaborasi model pengukuran yang mereka susun dengan mudah. Feldt dkk. (1987) mengatakan bahwa popularitas koefiesien alpha lahir karena beberapa faktor, antara lain: (a) teknik komputasinya relatif mudah, karena hanya memerlukan informasi berupa varian butir dan skor total, (b) distribusi sampling sudah diketahui sehingga penentuan interval kepercayaan pada populasi sangat dimungkinkan. Koefisien alpha juga banyak dipakai pada banyak literatur karena merupakan estimator yang moderat dalam mengestimasi reliabilitas.

Setiap properti psikometris dikembangkan berdasarkan asumsi-asumsi tertentu. Ketika skala yang kita kembangkan tidak pas dengan asumsi tersebut, akibatnya muncul beberapa masalah dalam proses analisis atau hasil analisis menjadi menyesatkan. Sebagai contoh, mengestimasi properti konsistensi inernal pada skala yang mengukur domain yang luas atau yang memuat butir-butir yang heterogen akan menghasilkan nilai alpha yang rendah. Sebaliknya, mengestimasi reliabilitas konstruk melalui analisis faktor konfirmatori (CFA) pada skala yang mengukur domain yang sempit (*narrow*), yang ditunjukkan dengan butir yang memiliki interkorelasi sangat tinggi, akan memunculkan hasil estimasi yang tidak dapat diterima (*non-admissible*). Interkorelasi antar butir yang sangat tinggi menjadi kontra produktif dalam analisis faktor karena menyebabkan masalah multikolinieritas (Byrne, 2009). Akibatnya, muncul nilai statistik yang tidak logis, misalnya nilai bobot faktor terstandarisasi di atas 1 atau nilai varians yang memiliki arah negatif.

Tulisan ini akan memaparkan isu-isu terkait dengan penggunaan koefisien alpha dalam analisia butir pada alat ukur (skala) psikologi. Tema mengenai koefisien alpha, terutama keterbatasannya sudah sangat banyak ditulis, namun dalam kebanyakan

disajikan dalam Bahasa Inggris dan muncul di dalam jurnal-jurnal psikometri dan pengukuran yang kurang banyak diakses oleh peneliti yang tidak mendalami psikometri. Mengingat perlunya mensosialisasikan perkembangan mutakhir psikometri kepada kalangan peneliti di Indonesia, maka penulis menyajikannya kembali.

Konsistensi Internal vs. Unidimensionalitas

Konsistensi internal dan unidimensionalitas merupakan dua properti psikometris yang memiliki kemiripan namun berangkat dari filosofi yang sedikit berbeda dalam memandang sebuah pengukuran. Konsistensi internal menunjukkan derajat interkorelasi butir-butir skala, sedangkan unidimensionalitas menunjukkan bahwa butir-butir skala mengukur satu konstruk ukur. Dalam analisis butir, kedua properti ini banyak dipakai sebagai landasan untuk menyeleksi butir. Nilai konsistensi internal dapat dimaksimalkan dengan memilih butir-butir yang memiliki interkorelasi yang tinggi atau yang korelasi butir-total yang tinggi. Sementara itu, unidimensionalitas dapat dicapai dengan memilih butir-butir yang memiliki bobot faktor (*item loading*) yang tinggi. Wacana terkini menunjukkan bahwa secara umum konsistensi internal bukan menjadi prioritas dalam penyusunan skala. Hattie (1985) mengatakan bahwa poin penting dari penyusunan skala bukanlah konsistensi internal, melainkan undimensionalitas. Hal yang sama juga dikemukakan oleh beberapa ahli yang lain (Boyle, 1991; Clark & Watson, 1995; Miller, 1995).

Alwin (2007) membedakan dua pendekatan dalam penyusunan alat ukur, yaitu pendekatan pengukuran majemuk (*multiple measures*) dan indikator majemuk (*multiple indicators*). Pendekatan pengukuran majemuk mengasumsikan bahwa butir-butir di dalam skala merepresentasikan pengukuran terhadap satu konstruk ukur yang dilakukan berulang kali. Konsekuensinya, skala yang baik terlihat dari hasil pengukuran kedua memiliki kesamaan dengan pengukuran pertama dan seterusnya. Oleh karena antar butir merepresentasikan pengukuran yang berulang maka kesamaan yang dimaksud adalah kesamaan dalam hal rerata (presisi) dan variasi (skala) skor. Menurut pendekatan ini skala yang baik berisi butir-butir yang konsisten dalam mengukur, ditunjukkan dengan interkorelasi antar butir yang tinggi. Semakin tinggi nilai konsistensinya semakin menunjukkan pengukuran yang optimal. Konsistensi internal merupakan properti yang mewakili pendekatan pengukuran majemuk, yang biasa diestimasi dengan menggunakan teknik belah dua (split-half) atau menggunakan koefisien alpha.

Pendekatan indikator majemuk memiliki pandangan yang lebih moderat dibanding dengan pengukuran majemuk. Butir-butir dilihat sebagai sampel indikator dari populasi indikator yang memanifestasikan konstruk ukur. Oleh karena konstruk psikologi sangat kompleks, maka antara indikator satu dengan indikator lainnya tidak diasumsikan memiliki korelasi yang sangat tinggi. Yang dibutuhkan adalah asumsi bahwa butir-butir skala mengukur satu atribut yang sama (unidimensi). Teknik analisis faktor mewakili pendekatan ini karena dapat menghasilkan informasi seberapa jauh

butir-butir skala memanifestasikan konstruk secara optimal. Dengan menggunakan pendekatan indikator majemuk, maka tujuan analisis butir adalah untuk mendapatkan skala yang bersifat unidimensi daripada konsistensi internal tinggi (Clark & Watson, 1995).

Paradoks Atenuasi

Konsistensi internal yang tinggi ditunjukkan dengan tingginya interkorelasi antar butir. Konsep ini didasari pada asumsi bahwa jika butir-butir memiliki korelasi yang tinggi dengan konstruk target ukur, maka mereka akan memiliki korelasi yang tinggi juga satu sama lainnya. Sederhananya, korelasi antar butir yang tinggi dapat dicapai jika subjek memilih opsi respons "sangat sesuai" pada butir 1, korelasi antar butir akan menjadi tinggi jika subjek juga menjawab "sangat sesuai" juga pada butir 2 dan seterusnya. Kritik terhadap konsistensi internal muncul, upaya mengejar konsistensi internal yang sangat tinggi dinilai menjadi kontra produktif karena hanya akan menghasikan butir-butir yang saling tumpang tindih (*redundant*) (Boyle, 1991; Clark & Watson, 1995). Jika peneliti hanya mengejar untuk mendapatkan konsistensi internal tinggi tanpa memperhatikan konten butir-butir yang tersisa maka upaya ini dapat melemahkan validitas konten skala. Domain yang diukur oleh skala dapat menjadi lebih sempit dari desain awalnya karena di dalamnya hanya berisi butir-butir yang memiliki konten yang sama. Paparan ini menunjukkan adanya paradoks antara upaya meningkatkan reliabilitas (konsistensi internal) dan validitas: meningkatkan reliabilitas justru diikuti dengan menurunnya validitas, dan sebaliknya. Paradoks ini dinamakan dengan paradoks atenuasi (*attenuation paradox*) (Loevinger, 1954).

Paradoks ini memperlihatkan bagaimana seorang peneliti dapat menyusun skala yang memiliki reliabilitas tinggi akan tetapi tidak banyak informasi yang dihasilkan dari pengukuran tersebut. Clark dan Watson (1995) mencontohkan tiga butir yang mengukur kecemasan sosial, misalnya "*Saya merasa tidak nyaman ketika berada dalam sebuah kerumunan orang*", "*Saya biasanya merasa cemas ketika berada dalam acara sosial*" dan "*Tidak mudah bagi saya untuk bertemu dengan banyak orang*". Dilihat dari pernyataan yang tertulis, ketiganya memiliki kemiripan sehingga responden akan mendukung ketiga pernyataan tersebut atau sebaliknya, menolak ketiganya. Oleh karena kontennya tumpang tindih, maka tidak banyak informasi yang diungkap oleh ketiga butir tersebut. Butir-butir yang memiliki korelasi yang sangat tinggi bahkan korelasi sempurna justru tidak memberikan informasi yang baru.

Paradoks ini mengingatkan agar peneliti mempertimbangkan konten ukur ketika menyeleksi butir. Upaya memaksimalkan reliabilitas konsistensi internal perlu dibarengi dengan menjaga domain ukur tetap terjangkau oleh skala. Penulis merekomendasikan agar peneliti melakukan proses perakitan (*test assembling*) setelah analisi butir dilakukan. Analisis butir adalah upaya menggali informasi psikometris butir dan bukan hanya sekedar proses menyeleksi butir. Proses seleksi butir lebih dominan terjadi pada proses perakitan, karena di dalam proses ini pemilihan butir dilakukan tidak hanya

mempertimbangkan properti psikometris butir, melainkan keterwakilan domain ukur oleh butir, lama waktu pengadministrasian skala, panjang skala, atau tujuan pengukuran.

Asumsi yang Mendasari Koefisien Alpha

Sebuah properti psikometris selalu memiliki asumsi yang mendasari, demikian juga koefisien alpha. Sebelum mengulas lebih lanjut mengenai model pengukuran yang mendasari koefisien alpha, penulis akan menjelaskan empat model pengukuran dalam teori klasik, yaitu: model paralel, ekivalensi nilai tau, ekivalensi nilai tau essensial, dan konjenerik. Pemahaman terhadap model pengukuran ini sangat penting sebelum memilih koefisien atau indeks mana yang akan dipakai mengevaluasi pengukuran (Widhiarso, 2013). Ketidaktepatan dalam memilih model yang sesuai dengan skala yang dikembangkan akan menghasilkan informasi yang menyesatkan.

Keempat model tersebut dapat ditinjau berdasarkan dua pendekatan, yaitu pendekatan teori skor murni dan pendekatan analisis faktor. Menurut teori skor murni, dua statistik yang membedakan keempat model pengukuran di atas adalah (a) nilai harapan (*expected value*) yang merujuk pada parameter rerata skor murni yang diukur oleh butir/tes dan (b) nilai parameter varians (Traub, 1994). Menurut pendekatan analisis faktor, keempat model dapat dibedakan berdasarkan bobot faktor dan varians eror pengukuran.(Graham, 2006; Raykov, 2001). Kedua properti tersebut menjelaskan unit pengukuran atau skala dan presisi pengukuran. Nilai rerata menjelaskan presisi pengukuran, sedangkan nilai varians menjelaskan unit pengukuran atau skala ukur. Sebagai contoh, dua butir yang mengukur afek negatif yang menggunakan model Likert dengan lima opsi (0-4).

Tabel 1. Perbandingan Model Pengukuran Teori Klasik

Model Pengukuran	Teori Skor Murni		Analisis Faktor	
	Nilai Harapan	Varians	Bobot Faktor	Eror Pengukuran
Paralel	Sama	Sama	Sama	Sama
Ekivalensi-Tau	Sama	Bervariasi	Sama	Bervariasi
Ekivalensi-Tau Esensial	Bervariasi*)	Bervariasi	**)	**)
Konjenerik	Bervariasi	Bervariasi	Bervariasi	Bervariasi

Catatan. *) Nilai harapan antar butir mengikuti fungsi aditif, sehingga meski nilai harapannya adalah berbeda namun selisih nilai antar butir adalah sama. **) Dalam analisis faktor, adanya perbedaan aditif antar skor butir tidak mempengaruhi komputasi varians sehingga model tau ekivalen dan tau ekivalen esensial dinilai sama.

Butir pertama memuat pernyataan "*Saya terkadang merasa sedih*" sedangkan butir kedua memuat "*Saya hampir selalu merasa sedih*". Kedua butir cenderung memiliki distribusi skor yang sama pada populasi subjek yang diukur sehingga dapat dikatakan memiliki skala yang sama. Namun demikian, butir kedua memiliki intensitas yang kuat

(*strongly worded*) dibanding dengan butir kedua, yang menyebabkan keduanya memiliki presisi yang berbeda. Butir pertama akan menghasilkan rerata skor butir yang lebih tinggi dibanding dengan butir kedua karena peristiwa "terkadang" lebih sering terjadi atau dialami pada subjek dibanding dengan peristiwa "hampir selalu". Tabel 1 menunjukkan perbandingan keempat model pengukuran tersebut.

Dari Tabel 1 dapat disimpulkan bahwa model paralel membutuhkan asumsi yang ketat karena mengasumsikan bahwa skala dan presisi pengukuran semua butir adalah sama. Model ekivalensi-tau dan ekivalensi-tau esensial sedikit moderat, mengasumsikan skala sama akan tetapi memiliki presisi yang bervariasi. Model konjenerik memiliki asumsi yang paling moderat, memperbolehkan presisi dan skala pengukuran yang bervariasi (Graham, 2006). Dari keempat model pengukuran di muka, koefisien alpha mengasumsikan bahwa skala yang dievaluasi mengikuti model ekivalensi-tau esensial yang ditunjukkan dengan korelasi yang tinggi antara skor murni komponen satu dengan lainnya, serta varian skor antar komponennya adalah setara (Feldt & Ankenmann, 1999). Persyaratan tersebut kemudian ditambahkan lagi dengan berupa kovarian skor tampak yang setara. Asumsi mengenai kesetaraan ini menunjukkan bahwa koefisien alpha akan menghasilkan nilai yang tinggi ketika butir-butir yang diuji bersifat homogen. Sebagai catatan, Schmitt (1996) mengatakan bahwa setelah mengestimasi reliabilitas alpha, peneliti diharapkan untuk melaporkan presisi nilai alpha atau eror standar alpha (*standard error of alpha*) agar peneliti lain dapat mengetahui kemungkinan adanya bias di dalam koefisien tersebut. Azwar (2007) mengatakan bahwa koefisien alpha yang sama memungkinkan memiliki presisi yang berbeda karena keduanya dapat saja memiliki eror standar yang berbeda.

Dapat disimpulkan bahwa, jika peneliti hendak menggunakan pendekatan konsistensi internal melalui koefisien alpha dalam mengevaluasi skala maka diharapkan mempertimbangkan hal-hal berikut.

a) Skala harus mengukur atribut tunggal (unidimensi).
b) Butir-butir skala memiliki nilai relevansi dan intensitas yang setara. Di sisi lain, konten ukur ditulis dalam pernyataan yang memiliki penekanan sama, sehingga diharapkan tidak ada pernyataan yang memiliki penekanan kuat (e.g. "*Saya membenci…*") bercampur dengan penekanan lemah (e.g. "*Saya kurang suka..*") dalam satu skala.
c) Sampel yang ditargetkan untuk diukur memiliki level konstruk yang sama. Misalnya, jika mengembangkan skala yang mengukur depresi, maka peneliti perlu menentukan level depresi mana yang akan diungkap oleh skala, level rendah, sedang atau menengah. Mencampurkan butir yang mengukur depresi level rendah (e.g. sulit tidur) dan tinggi (e.g. bunuh diri) dalam satu skala cenderung akan mendapatkan konsistensi internal yang rendah. Dengan demikian dapat diketahui skala berisi butir dengan intensitas (i.e tingkat kesulitan) beragam seperti yang dikembangkan oleh skala interval tampak setara dari Thurstone kurang tepat dievaluasi dengan menggunakan pendekatan konsistensi internal.

Intinya, adanya variasi dalam skala baik menyangkut konten, format, karakteristik butir dan penulisan pernyataan memungkinkan asumsi ekivalensi-tau menjadi terlanggar. Oleh karena itu, jika peneliti menggunakan konsistensi internal, diharapkan menelaah lebih mendalam adanya variasi-variasi tersebut.

Pengujian Ketepatan Model Pengukuran

Evaluasi model pengukuran yang dilakukan oleh penulis dengan menggunakan kedua pendekatan untuk mengevaluasi beberapa skala psikologi menemukan bahwa skala psikologi cenderung memenuhi asumsi konjenerik dibanding dengan asumsi paralel atau ekivalensi-tau. Pada penelitian pertama (Widhiarso, 2012), penulis menguji kesetaraan rerata skor butir yang dilakukan dengan menggunakan uji Hotelling-T2, dengan hipotesis nihil bahwa semua butir skala memiliki rerata yang sama. Uji kesetaraan varians dilakukan dengan menggunakan uji kai-kuadrat dengan hipotesis nihil bahwa semua butir dalam skala memiliki varians butir dan eror yang setara. Hasil analisis (Tabel 2) menunjukkan bahwa butir-butir di dalam skala yang diuji, cenderung memiliki rerata, varians dan varians eror yang bervariasi.

Tabel 2. Hasil pengujian kesetaraan rerata, varians dan varians eror butir

Pengukuran	Uji Kesetaraan Rerata		Uji Kesetaraan Varians	
	T2	db	χ^2	db
BDI Faktor 1	1057.355		675.409	19
BDI Faktor 2	2909.784	3	1215.921	8
BDI Faktor 3	47.876	3	527.742	8
BDI Faktor 4	412.256	7	1083.693	34
BDI Faktor 5	1868.608	4	538.239	13
Harga Diri	1253.587	9	3070.782	53

Catatan. Semua uji perbandingan menunjukkan hasil signifikan pada taraf 1% ($p<0.01$)

Hasil penelitian kedua (Widhiarso & Kožený, 2013) yang menggunakan pendekatan analisis faktor konfirmatori, menemukan bahwa model yang tepat untuk menjelaskan tiga skala yang diuji (BFI-44 dan harga diri) adalah model konjenerik. Perbandingan ketepatan model yang dilakukan dengan uji kai-kuadrat menghasilkan ketepatan model yang berbeda secara signifikan antara model yang diuji (Tabel 3). Secara umum, model pengukuran konjenerik mewakili ketiga skala yang diuji.

Temuan di atas menunjukkan bahwa skala psikologi yang telah dikembangkan selama ini cenderung memiliki butir-butir yang heterogen ketimbang butir-butir yang homogen. Meski sebuah skala mengacu pada konten yang sama, antara satu butir dengan butir lainnya cenderung memiliki penekanan yang berbeda dalam mengukur. Beberapa hal yang mendukung munculnya heterogenitas butir di dalam skala psikologi

adalah sebagai berikut. Faktor-faktor yang dijelaskan di bawah ini juga menjadi penyebab skala psikologi menjadi bersifat multidimensi dan menurunkan nilai reliabilitas alpha (Widhiarso, 2009).

Tabel 3. Hasil pengujian Model Pengukuran dalam Perspektif Analisis Faktor

Pengukuran	Paralel vs. Ekivalensi-Tau	Paralel vs. Konjenerik	Ekivalensi vs. Konjenerik
BFI-44			
Faktor 1	229.156 (7)	251.789 (14)	22.633 (7)
Faktor 2	243.097 (8)	268.742 (16)	25.645 (8)
Faktor 3	241.005 (8)	322.241 (16)	81.236 (8)
Faktor 4	90.961 (7)	534.708 (14)	443.747 (7)
Faktor 5	639.719 (9)	1057.733 (19)	418.014 (10)
Harga diri	716.94 (9)	2260.54 (18)	1543.6 (9)

Catatan. Semua uji perbandingan menunjukkan hasil signifikan pada taraf 1% ($p<0.01$)

a). Karakteristik konstrak psikologi yang kompleks. Skala harga diri dari Coopersmith adalah skala ditujukan untuk mengukur konstruk tunggal, namun hasil analisis pada data yang didapatkan di lapangan menunjukkan bahwa butir-butir di dalamnya bersifat heterogen (Aluja, Rolland, García, & Rossier, 2007).

b). Adanya penjabaran konstruk menjadi beberapa komponen atau aspek sebelum diterjemahkan menjadi indikator yang lebih operasional. Penyusunan skala psikologis seringkali diawali dari penurunan butir dari beberapa aspek keperilakukan, misalnya penyusunan skala efikasi diri yang diturunkan dari aspek atletik, akademik, dan kehidupan sosial (Czerniack, 2002). Poin yang ditekankan di sini adalah penjabaran konstruk ukur menjadi menjadi beberapa komponen cenderung menambah tingkat heterogenitas butir-butir skala (Streiner, 2003). Jika peneliti tidak mengkaji lebih mendalam mengenai hubungan antar aspek, maka butir-butir yang dihasilkan dari penjabaran aspek tersebut akan bersifat heterogen. Hal ini terjadi jika masing-masing aspek tidak memiliki kaitan. Akibatnya aspek tersebut menjadi bersifat independen antara satu dengan lainnya.

c). Jumlah butir yang dilibatkan di dalam skala. Drolet dan Morisson (2001) menunjukkan bahwa heterogenitas skala psikologi salah satunya dipengaruhi oleh jumlah butir yang dilibatkan. Jumlah butir yang terlalu banyak dapat menambah potensi munculnya varian-variaab yang tidak relevan dengan konstruk ukur, sehingga memunculkan dimensi baru dari dimensi yang ditetapkan awalnya.

d). Adanya arah pernyataan yang berbeda. Menambahkan butir yang memiliki arah negatif (*unfavorable*) selain butir yang positif (*favorable*) digunakan untuk mereduksi munculnya tanggapan responden yang monoton. Namun dengan penulisan yang kurang hati-hati, butir-butir di dalam skala akan menjadi heterogen dengan adanya butir berarah negatif. Analisis faktor yang dilakukan oleh Marsh (1996) terhadap

pengukuran harga diri menemukan bahwa butir-butir membentuk faktor berdasarkan arahnya. Butir berarah positif membentuk faktor 1 sedangkan sisanya (butir berarah negatif) membentuk faktor 2. Di sisi lain, nilai intensitas butir negatif lebih besar dibanding dengan butir positif. Hal ini dikarenakan secara kognitif, responden lebih mudah untuk menyetujui dibanding dengan menolak sebuah pernyataan yang diberikan kepadanya. Menurut Anderson (1965) pernyataan negatif memiliki bobot informasi yang lebih dibanding dengan pernyataan yang positif.

e). Konsep butir sebagai sampel indikator. Dalam skala psikologi, butir-butir merupakan sampel perilaku yang mewakili populasi perilaku konstruk ukur yang jumlahnya tidak terbatas. Dengan banyaknya indikator yang diungkap, maka tidak semua indikator keperilakuan dilibatkan dalam pengukuran. Oleh karena itu, ketika menulis sebuah butir, penyusun skala memiliki 'kekebasan' dalam menuangkan konsep teoritik menjadi butir-butir skala. Penyusun memiliki berbagai alternatif untuk memilih perilaku apa saja yang akan diturunkan menjadi butir. Adanya keragaman cara inilah yang menyebabkan butir-butir di dalam skala psikologi cenderung bersifat heterogen. Kondisi ini 'diperparah' dengan masuknya kata penanda (*modifier*) dalam kalimat pernyataan, misalnya penanda waktu ("*akhir-akhir ini"*, "*suatu saat*"), penanda subjek ("menurut teman saya", "menurut saya"), atau penanda lokasi ("*di sekolah*", "*di rumah*"). Masuknya penanda ini tujuannya adalah memperkaya informasi yang didapatkan dari proses pengukuran, namun konsekuensinya adalah butir-butir skala menjadi bersifat heterogen ketika hal itu tidak diterapkan secara konsisten pada semua butir.

Selain kelima faktor di atas, ada faktor lain yang juga menyebabkan munculnya sifat heterogenitas dalam pengukuran psikologi. Pertama, teori mengenai atribut yang hendak diukur kebanyakan tidak menyertakan secara mendetail operasionalisasi konsep teoritik pada level perilaku yang sangat empirik (Loevinger, 1954). Para penyusun skala sendiri lah yang banyak berperan dalam mengoperasionalisasikan konsep teoritik tersebut menjadi indikator keperilakuan yang nantinya menjadi butir skala . Kedua, tidak ada kaidah yang baku dalam menurunkan indikator menjadi kalimat pernyataan (Colton & Covert., 2007). Dengan tidak adanya kaidah baku penulisan butir, maka perilaku memukul sebagai indikator perilaku agresi dapat diturunkan menjadi butir "*Saya akan memukul orang yang menghina saya"* atau butir "*Saya akan memukuli orang yang menghina saya*". Jelas sekali bahwa meski sama-sama memukul namun kedua butir memiliki presisi ukur perilaku agresi yang berbeda.

Hasil penelitian yang dipaparkan di muka menunjukkan bahwa sebagian besar pengukuran dalam bidang psikologi cenderung bersifat heterogen daripada homogen. Namun demikian, masalah homogenitas butir di dalam skala tidak menjadi masalah ketika hal tersebut menjadi tujuan peneliti. Peneliti dapat sejak awal menetapkan bahwa butir di dalam skala yang dikembangkannya berisi butir-butir yang heterogen maupun homogen. Jika peneliti hendak mengembangkan skala yang berisi butir-butir heterogen, maka pendekatan konsistensi internal kurang tepat untuk dikenakan.

Sebagai alternatifnya, prosedur analisis faktor dapat dipakai untuk memberikan informasi mengenai properti psikometris butir.

Saran dan Rekomendasi

Penyusunan Alat Ukur

Rekomendasi yang dapat diberikan kepada para penyusun skala psikologi adalah memahami model pengukuran seperti apa yang akan diterapkan dalam skala yang disusun. Jika hendak mengembangkan skala yang memiliki konsistensi internal yang tinggi, penyusun dapat mengembangkan skala yang berisi butir-butir yang homogen dan redundan (*redundant*). Redundansi bukanlah sebuah hal yang buruk jika itu menjadi tujuan peneliti sejak awal (DeVellis, 2003). Redundansi menunjukkan bahwa butir-butir skala benar-benar memiliki kesamaan ukur serta konsisten dan bukan dikarenakan hal-hal yang kebetulan. Lebih lanjut DeVellis (2003) menambahkan bahwa skala membutuhkan adanya butir-butir redundan, terutama pada butir-butir yang diujicobakan daripada butir-butir di skala jadi. Di sisi lain, ada koefisien reliabilitas seperti Spearman-Brown dan alpha merupakan koefisien yang cocok diterapkan pada skala yang berisi butir-butir redundan atau skala yang memiliki sedikit variasi seperti yang dijelaskan di muka. Hal yang sama juga berlaku untuk skala yang mengukur atribut-atribut psikologi yang sangat spesifik, mengerucut dan memiliki jumlah butir yang sedikit. Misalnya, skala yang mengukur efikasi terhadap sebuah pelajaran matematika lebih memiliki domain yang mengerucut dibanding dengan skala yang mengukur efikasi secara umum.

Jika peneliti hendak mengembangkan skala yang mengukur atribut yang memiliki domain luas, memiliki banyak variasi dalam hal konten, format, atau intensitas ukur butir di dalam skala, maka model pengukuran konjenerik dapat menjadi acuan. Di sini, dimensionalitas pengukuran oleh skala menjadi prioritas (Clark & Watson, 1995). Dalam penyusunan skala, peneliti perlu menentukan apakah skala yang dikembangkan bersifat unidimensi atau multidimensi. Jika skala yang dikembangkan adalah skala yang bersifat unidimensi, peneliti perlu mengkaji aspek-aspek keperilakuan yang dijabarkan dari konstruk yang diukur. Aspek-aspek tersebut diharapkan memiliki keterkaitan yang secara singkat dapat dilihat dari tingginya level atribut individu pada aspek satu diikuti dengan tingginya level atribut individu pada aspek lainnya. Hal ini tentu saja dapat dijawab dengan menelaah dasar teori yang dipakai secara mendalam.

Dengan menggunakan model pengukuran konjenerik, peneliti tidak perlu melibatkan butir-butir yang redundan dan homogen. Redundansi dalam pendekatan ini malah akan menjadi hal kontraproduktif. Dengan adanya redundansi maka analisis faktor yang dilakukan dapat menghasilkan proses estimasi yang macet, model tidak fit atau munculnya rekomendasi indeks modifikasi untuk mengkorelasikan antar eror

pengukuran. Oleh karena sebelum uji coba pada sampel yang lebih besar (pra uji coba) butir-butir yang memiliki korelasi yang tinggi agar diperbaiki.

Dua saran yang dipaparkan di muka berdasar pada properti yang sama, yaitu redundansi. Mengembangkan skala yang mendasarkan pada model konsistensi internal menghendaki butir yang redundan, dan sebaliknya mengembangkan skala yang mengikuti model konjenerik tidak menghendaki butir yang redundan. Tantangan yang dihadapi kedua pendekatan cukup berbeda. Dalam model konsistensi internal, upaya untuk menyusun skala dengan butir-butir yang homogen dapat menjebak peneliti untuk menghasilkan skala validitas isinya rendah. Sebaliknya, tantangan dalam penyusunan skala dengan menggunakan model konjenerik adalah potensi munculnya dimensi-dimensi baru. Jika peneliti terlalu jauh keluar dari domain yang diukur, skala yang dihasilkan dapat menjadi tombak bermata dua, skala mengukur dua konstruk yang saling independen (Beuke, Fischer, & McDowall, 2003).

Analisis Data

Sebagai alternatif pendekatan analisis dengan menggunakan konsistensi internal, beberapa rekomendasi telah diberikan oleh para ahli. Berikut ini beberapa properti psikometris yang perlu diperhatikan ketika peneliti tidak menggunakan analisis butir yang berbasis model konsistensi internal.

a) *Korelasi antar butir*. Clark dan Watson (1995) menyarankan rerata agar rerata korelasi antar butir berada dalam interval 0,15 to 0,20 untuk skala yang mengukur konstruk yang relatif luas dan interval 0,4 hingga 0,5 untuk skala yang mengukur konstruk yang sempit.
b) *Korelasi butir-total*. Kline (1979, dikutip dari Panayides, 2013) menyarankan agar korelasi butir-total tidak melebihi 0,7 dengan tujuan untuk menjaga validitas konten atau luasnya domain ukur sesuai dengan desai awal.
c) *Membatasi nilai alpha*. Dengan nilai korelasi antar butir yang diharapkan cenderung moderat , maka nilai alpha yang sangat tinggi (di atas 0,9) tidak lagi menjadi prioritas. Beberapa penulis merekomendasikan nilai alpha yang ideal, diharapkan bergerak antara 0,8 hingga 0,9 (Clark & Watson, 1995; Panayides, 2013; Streiner, 2003). Menurut mereka, tingginya koefisien alpha dinilai tidak menunjukkan tingginya reliabilitas maupun kualitas skala. Nilai alpha yang terlalu tinggi menunjukkan adanya duplikasi dan replikasi konten yang diukur.Dalam buku yang banyak menjadi acuan meta kuliah psikometri, *Psychometric Theory* karangan Nunnally, penulis merekomendasikan nilai alpha sebesar 0,50 to 0,60 untuk penelitian awal, nilai alpha sebesar 0,80 untuk penelitian dasar, dan 0,90 sebagai nilai minimal yang ditoleransi untuk asesmen klinis. Dalam bukunya yang terbaru (Nunnally & Bernstein, 1984), penulis meningkatkan nilai alpha menjadi minimal menjadi 0,7.
d) *Analisis Faktor*. Boyle (1991) berpendapat bahwa untuk memaksimalkan luasnya domain pengukuran terhadap konstruk yang diukur, peneliti harus memilih butir dengan bobot (*factor loading*) yang tinggi pada faktor yang diukur tetapi pada saat

yang sama butir tersebut memiliki korelasi rendah hingga moderat dengan butir lainnya.

Rekomendasi lain yang dapat diberikan selanjutnya adalah menyesuaikan karakteristik skala yang dikembangkan dengan model pengukuran yang dipakai dalam menganalisis data. DeVellis (2003), menyarankan bahwa koefisien dapat dipakai jika skala yang dikembangkan mengukur konstruk yang sempit (*narrow*). Dua buah skala, Skala-A mengukur afek negatif sedangkan Skala-B mengukur kekhawatiran dan kecemasan akan cenderung menghasilkan nilai alpha yang berbeda. Skala-B cenderung akan menghasilkan nilai alpha yang lebih tinggi dibanding dengan Skala-A karena mengukur domain yang lebih sempit dibanding dengan Skala-B. Streiner (2003) mendukung pernyataan ini, dengan mengatakan bahwa koefisien alpha kurang cocok dipakai jika skala memiliki beberapa komponen (*multifaceted*).

Verifikasi Asumsi. Ketika peneliti belum yakin dengan model pengukuran seperti apa yang melekat pada skala yang dipakainya, peneliti dapat melakukan verifikasi asumsi. Peneliti perlu memverifikasi data hasil pengukuran sebelum mengindentifikasi properti psikometris berdasarkan formula yang tepat dengan karakteristik data dan tidak bergantung pada koefisien tertentu secara monoton. Sementara itu, Ferketich (1990) yang mengatakan bahwa seharusnya kajian dan pengujian reliabilitas tidak hanya terpaku pada satu koefisien saja melainkan juga melibatkan koefisien lain yang kemungkinan menggambarkan hasil yang lebih optimal.

Program bantu analisis seperti SPSS (2007) telah memfasilitasi upaya peneliti untuk memverifikasi asumsi data hasil pengukuran. Menu model paralel memverifikasi apakah butir memiliki kesetaraan dalam hal varians dan varians eror, sedangkan menu model paralel ketat (*strict paralel*) memverifikasi kedua asumsi dalam model paralel ditambah dengan asumsi kesetaraan nilai rerata. Verifikasi kesetaraan rerata dan varians skor butir tersebut dilaporkan pada bagian ketepatan (*goodness fit index*) melalui nilai kai-kuadrat. Nilai kai-kuadrat yang signifikan menunjukkan bahwa ada perbedaan yang signifikan antar rerata dan varians skor antar butir yang dianalisis. Selain itu ada uji Hotelling T2 yang dapat dipakai untuk menguji kesetaraan rerata saja. Verifikasi model pengukuran yang menggunakan pendekatan analisis faktor konfirmatori, dapat mengacu pada tulisan Widhiarso (2010).

*Penggunaan Analisis Faktor***.** Dari uraian di muka, terlihat bahwa korelasi antar butir menjadi satu properti psikometris yang penting untuk diidentifikasi dalam pengembangan skala. Masalahnya, jika skala yang dikembangkan memuat 30 butir yang diuji, maka akan dihasilkan 435 korelasi antar butir yang harus dikaji oleh peneliti. Untuk mengatasi kesulitan mengidentifikasi tabel yang sangat besar tersebut, sebagai gantinya kita dapat menggunakan analisis faktor. Penggunaan analisis faktor jarang dipakai, salah satunya adalah karena teknik ini membutuhkan ukuran sampel yang besar. Ukuran sampel yang direkomendasikan untuk menjalankan faktor analisis adalah minimal 200. Namun demikian ada beberapa pendapat yang cukup moderat (Gorsuch, 1983), mereka mengatakan bahwa analisis faktor boleh dijalankan dengan ukuran sampel minimal 100. Ahli lain (Mundfrom, Shaw, & Ke, 2005), menyatakan

bahwa penentuan ukuran sampel minimal tergantung dari nilai komunalitas, jumlah faktor yang diekstrak dan jumlah butir yang dianalisis. Berdasarkan studi simulasi yang mereka lakukan, analisis faktor dengan N=20 dapat dilakukan, asalkan faktor yang diekstrak adalah satu faktor dengan jumlah butir yang dianalisis adalah 7 butir. Berdasarkan tabel yang mereka susun, analisis faktor pada 21 butir yang menghasilkan 3 faktor telah memenuhi kaidah ketika menggunakan ukuran sampel antara 75 hingga 100 orang. Hasil penelitian menunjukkan bahwa ukuran sampel yang dibutuhkan dalam melakukan analisisi faktor relatif mudah dipenuhi sehingga penggunaan analisis faktor dalam penyusunan skala sangat ditekankan oleh para ahli.

Berikut ini contoh singkat tahap-tahap aplikasi analisis faktor dalam penyusunan skala yang dituliskan Clark dan Watson (1995) pada penyusunan alat ukur unidimensi. Langkah pertama, analisis faktor dilakukan dengan menetapkan sedikit faktor, misalnya empat atau lima faktor tanpa rotasi. Berikut ini catatan dalam mempertahankan atau mengugurkan butirnya.

a) Butir yang memiliki bobot yang rendah (di bawah 0,35) pada faktor atau komponen pertama,memiliki korelasi yang rendah dengan butir-butir lainnya. Butir ini dipertimbangkan untuk digugurkan.
b) Butir yang memiliki bobot lebih tinggi selain pada faktor pertama juga dipertimbangkan untuk digugurkan.
c) Butir yang memiliki bobot tinggi (di bawah 0,35) pada faktor pertama dan memiliki bobot rendah pada faktor yang lain, dipertimbangkan untuk dipertahankan

Peneliti juga dapat mencoba berbagai properti lain misalnya dengan mengurangi jumlah faktor yang diekstrak, mengganti teknik estimasi yang dipakai atau menambahkan rotasi faktor dalam analisis. Hasil yang memberikan informasi yang memiliki dukungan teori yang dipakai dalam penyusunan skala, ditetapkan sebagai hasil analisis akhir. Penggunaan prosedur analisis faktor dalam menyusun skala dapat merujuk pada tulisan Floyd dan Widaman (1995)

Langkah ini dapat dipakai sama halnya dengan proses seleksi butir yang biasa dilakukan, yaitu berdasarkan korelasi butir-total. Bedanya, kalau korelasi butir-total menggunakan skor tampak sebagai kriterianya sedangkan analisis faktor menggunakan skor faktor sebagai kriterianya. Skor faktor merupakan hasil estimasi terhadap skor murni yang sudah terbebas dari unsur eror pengukuran. Hasil analisis faktor dengan cara seperti ini akan menghasilkan butir-butir yang mengukur atribut tunggal (unidimensi). Analisis faktor selain bertujuan untuk melihat dimensionalitas pengukuran (unidimensional vs. multidimensinal) juga mengeksplorasi ragam presisi dan skala ukur tiap butir yang terlihat melalui bobot faktor tiap butir.

Penutup

Tulisan ini memaparkan keterbatasan penggunaan koefisien alpha sebagai properti psikometris pengukuran dalam skala psikologi karena butir-butir dalam skala psikologi cenderung bersifat heterogen. Saran praktis yang dapat diberikan kepada peneliti

adalah agar menyesuaikan analisis butir sesuai dengan karakteristik skala yang disusun. Hal ini dapat juga berlaku sebaliknya, agar peneliti menyusun skala yang sesuai dengan model pengukuran yang telah ditetapkan. Dalam tulisan ini, karakteristik yang ditekankan adalah konsistensi internal dan homogenitas butir di dalam skala. Berbagai teknik alternatif telah dikembangkan oleh para ahli yang berguna untuk mengevaluasi properti psikometris skala yang didusun. Selain reliabilitas alpha, berbagai koefisien reliabilitas dapat dipakai sebagai alternatif, misalnya reliabilitas komposit (Raykov, 1997). Pendekatan berbasis model Rasch juga menjadi alternatif ketika peneliti mengembangkan skala yang berisi butir-butir yang heterogen. Seperti halnya, skala inteval tampak setara dari Thurstone atau Skala Guttman model ini justru menekankan pada variasi butir yang dinyatakan dalam lokasi butir (*item location*). Dalam pemodelan Rasch, lokasi butir dalam skala psikologi diharapkan bervariasi (Bond & Fox, 2001).

Daftar Pustaka

Aluja, A., Rolland, J.-P., García, L. F., & Rossier, J. (2007). Dimensionality of the Rosenberg Self-Esteem Scale and Its Relationships With the Three-and the Five-Factor Personality Models. *Journal of Personality Assessment, 88*(2), 246-249. doi: 10.1080/00223890701268116

Alwin, D. F. (2007). *Margins of error : A study of reliability in survey measurement.* Hoboken, NJ: A John Wiley & Sons, Inc.

Anderson, N. H. (1965). Averaging versus adding as a stimulus-combination rule in impression formation. *Journal of Experimental Psychology: General, 70*(4), 394-400. doi: 10.1037/h0022280

Azwar, S. (2007). *Dasar-dasar psikometri.* Yogyakarta: Pustaka Pelajar.

Beuke, C. J., Fischer, R., & McDowall, J. (2003). Anxiety and depression: Why and how to measure their separate effects. *Clinical Psychology Review, 23*(6), 831-848. doi: http://dx.doi.org/10.1016/S0272-7358(03)00074-6

Bond, T. G., & Fox, C. M. (2001). *Applying the Rasch Model. Fundamental Measurement in the Human Sciences*. Mahwah NJ: Lawrence Erlbaum.

Boyle, G. (1991). Does item homogeneity indicate internal consistency or item redundancy in psychometric scales? *Personality and Individual Differences, 12*(3), 291-294.

Byrne, B. M. (2009). *Structural equation modeling with Amos: Basic concepts, applications, and programming (2nd ed.).* New York: Taylor & Francis.

Clark, L. A., & Watson, D. (1995). Constructing Validity : Basic Issues in Objective Scale Development The Centrality of Psychological Measurement. *Psychological Assessment, 7*(3), 309-319. doi: 10.1037/1040-3590.7.3.309

Colton, D., & Covert., R. W. (2007). *Designing and constructing instruments for social research and evaluation*. San Francisco, CA: John Wiley & Sons, Inc.

Cronbach, L. J. (1951). Coefficient alpha and the internal structure of tests. *Psychometrika, 16*(3), 297-334. doi: 10.1007/bf02310555

DeVellis, R. F. (2003). *Scale development: Theory and applications*. Newbury Park: SAGE Publications, Inc.

Drolet, A. L., & Morrison, D. G. (2001). Do We Really Need Multiple-Item Measures in Service Research? *Journal of Service Research, 3*(3), 196-204. doi: 10.1177/109467050133001

Feldt, L. S., & Ankenmann, R. D. (1999). Determining sample size for a test of the equality of alpha coefficients when the number of part-tests is small. *Psychological Methods, 4*(4), 366-377. doi: 10.1037/1082-989X.4.4.366

Feldt, L. S., Woodruff, D. J., & Salih, F. A. (1987). Statistical Inference for Coefficient Alpha. *Applied Psychological Measurement, 11*(1), 93-103. doi: 10.1177/014662168701100107

Ferketich, S. (1990). Focus on Psychometrics Internal Consistency Estimates of Reliability. *Researching Nursing & Health, 13*, 437-440.

Floyd, F. J., & Widaman, K. F. (1995). Factor analysis in the development and refinement of clinical assessment instruments. *Psychological Assessment, 7*, 286—299. doi: 10.1037/1040-3590.7.3.286

Gorsuch, R. L. (1983). *Factor Analysis, 2nd edition*. Hillsdale, NJ: Erlbaum.

Graham, J. M. (2006). Congeneric and (Essentially) Tau-Equivalent Estimates of Score Reliability: What They Are and How to Use Them. *Educational and Psychological Measurement, 66*(6), 930-944.

Hattie, J. (1985). Methodology review: Assessing unidimensionality of tests and items. *Applied Psychological Measurement, 9*(2), 139-164. doi: 10.1177/014662168500900204

Loevinger, J. (1954). The attenuation paradox in test theory. *Psychological Bulletin, 51*(493-504).

Marsh, H. W. (1996). Positive and negative global self-esteem: A substantively meaningful distinction or artifactors? *Journal of Personality and Social Psychology, 70*(4), 810-819. doi: 10.1037/0022-3514.70.4.810

Miller, M. B. (1995). Coefficient alpha: A basic introduction from the perspectives of classical test theory and structural equation modeling. *Structural Equation Modeling: A Multidisciplinary Journal, 2*(3), 255-273. doi: 10.1080/10705519509540013

Mundfrom, D. J., Shaw, D. G., & Ke, T. L. (2005). Minimum Sample Size Recommendations for Conducting Factor Analyses. *International Journal of Testing, 5*(2), 159-168. doi: 10.1207/s15327574ijt0502_4

Nunnally, J. C., & Bernstein, I. H. (1984). *Psychometric theory*. New York: McGraw Hill Inc.

Panayides, P. (2013). Coefficient alpha: Interpret with caution. *Europe's Journal of Psychology, 9*(4), 687-696. doi: 10.5964/ejop.v9i4.653

Raykov, T. (1997). Estimation of Composite Reliability for Congeneric Measures. *Applied Psychological Measurement, 21*(2), 173-184. doi: 10.1177/01466216970212006

Raykov, T. (2001). Bias of Coefficient for Fixed Congeneric Measures with Correlated Errors. *Applied Psychological Measurement, 25*(1), 69-76. doi: 10.1177/01466216010251005

Schmitt, N. (1996). Uses and abuses of coefficient alpha. *Psychological Assessment, 8*(4), 350-353.

Socan, G. (2000). Assessment of Reliability when Test Items are not Essentially t-Equivalent. In A. Ferligoj & A. Mrvar (Eds.), *Developments in Survey Methodology* Ljubljana: FDV.

SPSS Inc. (2007). *SPSS Base 16.0 User's Guide*. Chicago, IL: SPSS Inc.

Streiner, D. L. (2003). Starting at the Beginning: An Introduction to Coefficient Alpha and Internal Consistency. *Journal of Personality Assessment, 80*(1), 99-103. doi: 10.1207/s15327752jpa8001_18

Traub, R. E. (1994). *Reliability for the social sciences: theory and applications*: Sage.

Widhiarso, W. (2009). Koefisien reliabilitas pada pengukuran kepribadian yang bersifat multidimensi. *Psikobuana, 1*(1), 39 - 48.

Widhiarso, W. (2010). *Aplikasi Pemodelan Persamaan Struktural dalam Pengujian Model Pengukuran Teori Klasik. Makalah.* . Paper presented at the Temu Ilmiah Nasional dan Kongres XI HIMPSI, 18-20 Maret 2010 Surakarta Jawa Tengah.

Widhiarso, W. (2012). Pengujian Kesetaraan Presisi dan Skala Ukur Butir-butir pada Skala Psikologi. *Jurnal Kawistara, 2*(1), 87-96.

Widhiarso, W., & Kožený, J. (2013). Applying confirmatory factor analysis to evaluate the fit of measurement models in education and psychological assessment. *Journal of Educational Sciences and Psychology, 3*(2), 78 - 88